



THE METHOD OF AUTOMATED FORMATION OF THE SEMANTIC DATABASE MODEL OF THE DIALOG SYSTEM

Inteasar Yaseen Khudhair, Yasir Ali Matnee and Aqeel Thamer Jawad

Department of Computer Science, Basic Education College,
University of Diyala, IRAQ

ABSTRACT

The paper deals with the problem of intellectual database content analysis for creating a semantic database model. Voice assistants are created to simplify interactions with mobile devices such as smartphones and tablets. A text is a system of analogue of this approach. As a result, user can interact with software using natural language. User interface is a set of software solutions that helps to search, review, obtain and process information from a database that is external storage. The natural language interface is the sort of user interface that accepts and processes the natural language queries. This interface can also be used in natural language. A semantic database .This model includes interconnections and internal structure of a database. Manual creating of a semantic database model .The system of semantic database mode. The proposed method uses a set of approaches to an automated semantic model. An object domain thesaurus helps to define semantics and solve the problem of polysemy in text processing Patterns helps to extract interconnections in a database. The analysis of the field of the field. Locale indication allows decreasing the time for a database content analysis. The system of semantic database model .The proposed method uses a set of approaches to an automated semantic model. An object domain thesaurus helps to define semantics and solve the problem of polysemy in text processing. Patterns helps to extract interconnections in a database .The analysis of the field of the field. Locale indication allows decreasing the time for a database content analysis. The system of semantic database model. The proposed method uses a set of approaches to an automated semantic model. An object domain thesaurus helps to define semantics and solve the problem of polysemy in text processing. Patterns help to extract interconnections in a database.

Key words: Natural Language Processing, Semantic Model, User Interface, Database, Dialog System.

Cite this Article: Inteasar Yaseen Khudhair, Yasir Ali Matnee and Aqeel Thamer Jawad, The Method of Automated Formation of The Semantic Database Model of The Dialog System, International Journal of Civil Engineering and Technology, 9(7), 2018, pp. 1117–1122.
<http://www.iaeme.com/ijciyet/issues.asp?JType=IJCIET&VType=9&IType=7>

1. INTRODUCTION

Currently, there is a large number of operating systems and application software. Each of the programs has a specific interface of interaction, which leads to an increase in the time required for training the user to work with the system. To simplify the work with programs on mobile devices, such as smartphones and tablets, voice assistants are actively introduced. By analogy with the voice assistant, it is possible to introduce an interactive textual system, which is an interactive software system of human-machine interaction, where the user can form queries in the natural language. Thus, the user interacts with the software system of using a more familiar natural language [1].

The user interface is a system of software solutions that implement search, view, receive and process information from external storage - a database. The natural language interface is a kind of user interface that accepts and processes queries in the natural language and can use the natural language to output the information to the user [2]. There are various approaches to the development of natural language interfaces. For example, the NaLIR [3] project is based on the creation of a dependency tree, as well as the use of heuristics and rules in the process of parsing a natural language query, while the Sqlizer [4] uses machine learning methods.

The semantic model is one of the most important components of the dialogue system. This model describes the relationships and internal structure of the database [5]. Thus, the semantic model of the database can be used to solve problems related to the resolution of the ambiguity of natural language [6].

The process of processing a custom query in a natural language consists of the sequential execution of morphological, syntactic and semantic analysis [6]. Morphological and morphemic analyzes of the user's request are carried out first. Within the framework of the morphological analysis, the case, declination, and part of speech are determined. In morphemic analysis, each word is divided into separate morphemes: prefix, root, suffix, ending.

During the syntactic analysis, syntactic links within the sentence are singled out - the main and secondary members of the sentence, the type of the sentence. At this stage, the syntactic and lexical rules of the analyzed language are used, as well as the information obtained at the stage of morphological analysis [7].

The next step in the processing of a natural language query is the construction of a semantic representation. A semantic representation of the user's natural language query is built on the basis of the semantic model of the database.

2. FORMATION OF THE SEMANTIC MODEL OF THE DATABASE

The semantic model describes the entities, information about which is contained in the database. Also, the model includes the relationships between entities that are similar to those in the entity-relationship diagrams (ER diagrams). When developing a dialog system, the developer often faces the task of implementing a natural language user interface to existing and filled-in database data. In this case, the manual formation of the semantic database model leads to a significant increase in the amount of time and labor and, as a consequence, the cost of developing the software system.

To solve the problem, it is required to develop a mechanism that helps in the automated mode to form a semantic model of the existing database. The formation of the semantic model takes place in the following sequence (Figure 1):

- the names of all database tables are extracted;
- for each table, all fields are extracted;
- based on the names of the tables and their fields, the assumption is made about the contents, the internal links are determined;
- in the manual mode, clarification and expansion of information obtained at the previous stage is possible.

One of the problems preventing unambiguous correct determination of the contents of the database is the ambiguity of the natural language. Fields of tables with the same contents can have many different names depending on who asked them.

3. DEFINITION OF INTERNAL DATABASE CONNECTIONS

To determine the internal relationships between tables, it is possible to apply a number of patterns based on the naming of tables and DB fields. One of these patterns is the foreign key connection and the database table of the type `id_ [tableName]` or `[tableName] _id`, where `[tableName]` is the name of the database table to which the foreign key is associated.

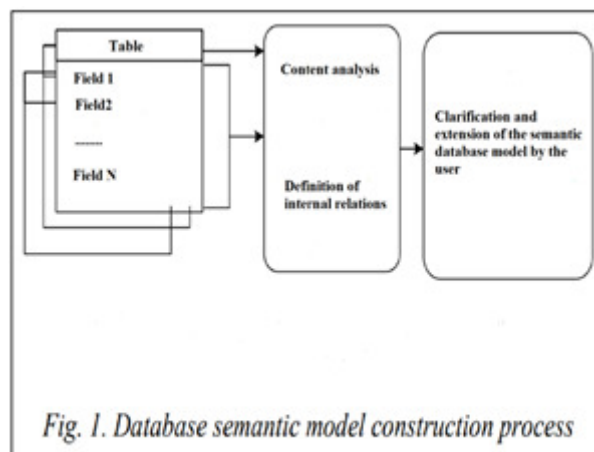
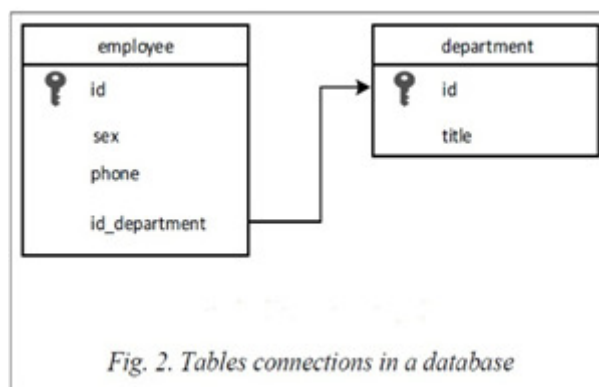


Fig. 1. Database semantic model construction process

Let us consider a concrete example (Fig. 2). There is an employee table - a relationship variable R2 and a department table - a variable of the ratio R1. In this case, the `id_department` field of the employee table is a foreign key (FK) whose values must match the values of the potential key (CK) of the R1 ratio variable. In the role of such a potential key is the primary key - the `id` field of the department table. The following conditions are fulfilled:

- in the variable of the relation R1 there is a potential key CK, such that FK and CK coincide up to the renaming of attributes;
- at any time, each FK value in the current R2 value is identical to the value of CK in some tuple at the current R1 value.

Thus, with the use of the pattern of the type `id_ [tableName]`, the relationship between the parent R1 (department) and the child R2 (employee) relations is found.



4. DEFINITION OF SEMANTICS AND TYPE OF CONNECTION BETWEEN DATABASE ENTITIES USING THESAURUS

One of the solutions to the problem of determining the semantics of data stored in one or another field of the database table is the application of the domain thesaurus. The thesaurus in general form is a dictionary, which includes the concepts, definitions and terms of a special field of knowledge [8]. Within this area of knowledge, information is stored in the database. The thesaurus can also include semantic relationships between lexical units, such as synonyms, antonyms, and the like. The thesaurus makes it possible to reveal meaning not only by definition, but also by relating the word to other concepts and their groups.

The thesaurus must be formed in advance. As a similar thesaurus, general solutions can be used, for example EuroWordNet [9], or own thesaurus. Own thesaurus can be independently formed automatically, for example, based on project technical documentation or other metadata.

In the above example, a relationship is found between the employee and the department located in different database tables. At the same time, the type of connection between the selected entities remains unknown. This problem can be solved if the user is given the opportunity to specify the type of communication manually after the automatic determination of entities.

As an alternative approach, a user thesaurus is created, which includes the types of relationships between entities stored in the database. Such a thesaurus is developed within the framework of an organization engaged in maintaining and filling the database. In this case, the connection between the essences employee and department is represented in the form of the following triplet: <employee, works in, department>.

In addition, by applying information from general thesauri, it is possible to determine the semantics of entities stored in the considered tables of the database. The result is the ratio given by the following triple: <employee, works in, department>.

5. DETERMINATION OF SEMANTICS BASED ON CONTENT ANALYSIS

One way to determine the semantics of the field in automatic mode is to analyze the contents of this field. Another approach to analyzing the contents of the database field is to use various text boxes. For example, in this way, it is possible to confirm the hypothesis of finding a ligature Surname-Name-Patronymic with the use of the corresponding text corpus.

As an additional opportunity to improve the quality of data analysis, it is possible to specify the data locale before the analysis of the contents of the database. This will also reduce the processing time, since in this case the rules, thesauruses and text boxes specific for the specified languages will be used.

6. RESULTS OF THE EXPERIMENTAL STUDY

An experimental investigation of the proposed method of automated formation of a semantic database model of a dialog system is carried out. For the experiment, a dialog system was created for which a semantic model was generated in an automated mode. Further, several volunteers in natural language formed questions to the experimental dialogue system. Thus formed $|Drel| = 130$ questions.

Based on the results of interaction with the dialog system, the correctness of the generated SQL-queries to the database and the relevance of the received response were evaluated manually. Syntactically correctly formed were $|Dretr| = 84$ SQL queries. In this case, the relevant answer was obtained in $|Drel \cap Dretr| = 75$ cases. Subsequent analysis of situations in which it was not possible to build a correct SQL query or obtain a relevant response showed that using a custom thesaurus containing information on abbreviations and abbreviations would improve the result obtained. As a result, the extraction accuracy is $Pr = |Drel \cap Dretr| / |Dretr| = 0.89$, completeness $Re = |Drel \cap Dretr| / |Drel| = 0.58$, the combined F-metric $2PrRe / (Pr + Re) = 0.70$.

7. CONCLUSION

The semantic model of the database is an important component of the dialogue system. Forming a semantic database model manually leads to a significant increase in time and labor, the cost of developing a software system.

The paper suggests a number of approaches that allow in an automated mode to form a semantic model of an existing database. Using the domain thesaurus makes it possible to define semantics, largely solving the problem of polysemy when interpreting the text. The use of patterns makes it possible to identify connections within the database, the intelligent analysis of the contents of the database fields - to determine the nature of the stored data, and the indication of the locale - to shorten the time required for analyzing the contents of the database. The proposed approach eliminates the shortcomings of existing solutions, since it assumes that there is information about the internal structure of the database, which helps to build a more accurate SQL query.

REFERENCES

- [1] Androutopoulos, Ion, Graeme D. Ritchie, and Peter Thanisch. 1995 Natural language interfaces to databases—an introduction. *Natural language engineering* 1.01: 29- 81.
- [2] Boonlit Adipat , Dongsong Zhang , Lina Zhou, 2011 The effects of tree-view based presentation adaptation on mobile web browsing, *MIS Quarterly*, v.35 n.1, p.99-122, March
- [3] N. Nihalani, S. Silakari, and M. Motwani, 2011 “Natural language Interface for Database: A Brief review” in *International Journal of Computer Science Issues (IJCSI)*, 8(2).
- [4] Jeffrey P, Richard S, Christopher D, 2014, GloVe: Global Vectors for Word Representation, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- [5] Giordani A., Moschitti A., 2009, Semantic mapping between natural language questions and SQL queries via syntactic pairing. *Intern. Conf. on Application of Natural Language to Inform. Syst.* Springer Berlin Heidelberg, pp. 207-221.
- [6] Ahmad, R., Abid, M., Ali, R.: 2009, Efficient Transformation of Natural Language Query to SQL. In: *Proc. 2nd Conference on Language and Technology*, pp. 53–60.
- [7] Bhadgale AM, Gavvas SR, Goyal PR, 2013, Natural language to SQL conversion system. *IJCSEITR.* ,vol. 3, i. 2, pp. 161-166.

- [8] Hamilton N. 2003, the mechanics of a deep net metasearch engine // Proc. of the 12th Intern. World Wide Web Conf., Budapest, Hungary.
- [9] Vossen P. EuroWordNet: a multilingual database for information retrieval. Proc. DELOS Workshop on Cross-language Information Retrieval, 1997, pp. 5-7.
- [10] Resnick P. Internet message format. 2008. URL: <http://www.ietf.org/rfc/rfc5322.txt> (date of circulation: January 14, 2018).